EDICIONES
COMPLUTENSE

Revista Electrónica Complutense de **Investigación en Educación Musical**
e-ISSN: 1698-7454

ARTÍCULOS

Investigating the impact of Music-Speech Integration Pedagogy (MSIP) on children's music aptitude: A registered report

Siu-Hang Kong
Department of Early Childhood Education, The Education University of Hong Kong (Hong Kong)
Chiu Ming Ming
Analytics/Assessment Research Centre, The Education University of Hong Kong (Hong Kong)
Department of Special Education and Counselling, The Education University of Hong Kong (Hong Kong)
William Choi
Speech and Music Perception Laboratory, The University of Hong Kong (Hong Kong)
Centre for Advancement in Inclusive and Special Education, The University of Hong Kong (Hong Kong)
Academic Unit of Human Communication, Learning, and Development, The University of Hong Kong (Hong Kong)
Alfredo Bautista
Department of Early Childhood Education, The Education University of Hong Kong (Hong Kong)

https://www.doi.org/10.5209/reciem.100735 Received: 04/02/2025 • Accepted: 15/11/2025 • Published: 27/03/2026

ENG **Abstract:** This Stage 1 Registered Report describes the design and analysis plan for a randomised controlled trial (RCT) evaluating the Music-Speech Integration Pedagogy (MSIP), a small-group programme for boosting kindergartners' tonal-rhythmic sensitivity and transfer to phonological awareness. As music and speech share acoustic attributes and overlapping neural systems, training that strengthens tonal-rhythmic processing may enhance phonological awareness. We will enroll 40 Cantonese-speaking children aged 4–5 years in Hong Kong. Half will learn with MSIP and half proceed as usual. We will primarily assess pre–post changes in music and speech ability with the Abbreviated Primary Measures of Music Audiation (aPMMA-T/R) and in English Phonological Awareness (EPA) test. Also, these children will complete an absolute pitch test (APT). We will analyse the RCT data with *multilevel, multivariate outcome differences-in-differences models*, supplemented with sensitivity analyses and statistical matching methods. Our findings will show whether MSIP is feasible and its efficacy for strengthening foundational auditory skills and prosody perception. Furthermore, it will clarify music-speech links in a tone-language context.
Keywords: music intervention; preschool; kindergarten; tonal-rhythmic sensitivity; phonological awareness; randomised controlled trial; Cantonese.

Table of Contents:

How to cite: Kong, S.-H, Chiu, M. M., Choi, W. & Bautista, A. (2026). Investigating the impact of Music-Speech Integration Pedagogy (MSIP) on children's music aptitude: A registered report. *RECIEM. Revista Electrónica Complutense en Educación Musical, 23*(1), 153-165, https://www.doi.org/10.5209/reciem.100735

1. Introduction

Music and speech share structural and functional commonalities, engaging overlapping neural networks involved in auditory, rhythmic, and prosodic processing (e.g., Broca's and Wernicke's areas; Patel, 2008, 2011). Both rely on pitch, rhythm, and timing cues to convey meaning and emotion. Evidence shows that training in one domain can enhance performance in the other. For instance, musical training has been found to enhance phonological awareness (PA) and speech perception (Moreno et al., 2015; Patel, 2008), while language experience, particularly in tonal languages, can sharpen musical pitch perception (Bidelman et al., 2013; Creel et al., 2023). Moreover, Cantonese speakers with higher musical pitch and rhythm perception show better English stress perception (Choi, 2022). These findings suggest a reciprocal reinforcement effect between music and speech processing.

Despite this well-established overlap, intervention studies examining transfer effects between music and speech have produced inconsistent results. Dumont et al. (2017), in their review of 46 studies, highlighted wide variability in intervention content, duration, and measurement approaches. Many programmes offered general music activities (e.g., singing, movement, instrument play) without systematically targeting the tonal-rhythmic dimensions most relevant to speech processing. Few interventions have purposefully applied language-related pedagogical methods to teach music, and none have explicitly leveraged the shared cognitive and neural underpinnings of music and speech to develop an integrated training programme aimed at improving both musical and linguistic skills.

Recent studies suggest that enhancing tonal-rhythmic sensitivity—the ability to perceive and organise pitch and time patterns—may be key to developing both musical aptitude and phonological awareness (Choi, 2022). Aural training focused on tonal–rhythmic accuracy has been shown to refine auditory discrimination skills and strengthen suprasegmental aspects of language perception, such as stress and intonation (Dunbar-Hall, 1991; Failoni, 1993; Lundsteen, 1971; Rivers, 1981). Likewise, rhythmic synchronisation and melodic contour tracking are central to both music and language processing. These converging findings imply that structured, pitch- and rhythm-based experiences could serve as a shared mechanism for improving Absolute Pitch (AP) and PA.

The present Stage 1 Registered Report introduces the Music–Speech Integration Pedagogy (MSIP)—a research-informed educational framework designed to refine children's auditory-perceptual skills by tapping into shared music–speech pathways in the brain. MSIP integrates listening, imitation, and Solfège-based tonal-rhythmic training to sharpen pitch and timing sensitivity. We hypothesise that targeted aural training will enhance tonal-rhythmic sensitivity, AP, and PA, thereby fostering the co-development of musical and linguistic abilities.

Registered reports prioritise the scientific merit of the research question and the rigour of the study design over the direction or significance of the results. In this model, editors and reviewers evaluate theoretical rationale and methodology before data collection, enhancing transparency, reproducibility, and methodological integrity while mitigating risks of bias and questionable research practices (Syed et al., 2023).

This registered report will (1) present the theoretical and empirical background linking music and speech; (2) specify the mechanistic pathways motivating MSIP; (3) describe the intervention design addressing limitations of prior studies; (4) outline the Stage 1 evaluation plan; and (5) discuss the study's expected contributions and limitations. In this study, tonal-rhythmic sensitivity refers to the ability to perceive and process pitch- and time-based information in a musical context. It is measured using the abbreviated Primary Measures of Music Audiation (aPMMA-T/R) and an Absolute Pitch (AP) test. Phonological awareness (PA) refers to the metalinguistic ability to consciously attend to and manipulate the sound structure of spoken language, encompassing syllables, phonemes, and suprasegmental prosody such as stress and intonation (Anthony & Lonigan, 2004; Goswami & Bryant, 1990). Because kindergarteners may have limited lexical knowledge, we focus on suprasegmental PA measures. By blending the strengths of music and language learning in a pedagogical framework, this study seeks to establish MSIP as an evidence-based approach for enhancing early musical and linguistic development.

2. Background: Shared acoustic mechanisms of music and speech

From a cognitive science and psychology of music perspective, the relationship between music and speech is understood as grounded in shared perceptual and neural mechanisms. Both domains rely on common acoustic dimensions—pitch and rhythm—which engage overlapping cognitive and neural processes (Tillmann, 2014; Patel, 2008).

Music and speech share functional and acoustic features, such as pitch and rhythm. Pitch is the sound frequency (*tonality*) of a musical note (Krumhansl & Cuddy, 2010), while rhythm is the timing of successive sounds. In speech, pitch can reflect speaker emotions and elicit them from listeners (Stolarski, 2015). Stolarski's (2015) study argued that a higher mean pitch and greater pitch variability tend to express and evoke positive emotions (such as happiness and joy). In contrast, flatter pitch contours, especially lower pitch variability, are commonly used to express and evoke sadness.

Rhythm and melody shape music and language development. Rhythm, the order of sounds across time, is crucial for music. Babies can hear rhythms to find spoken words, and thereby tell languages apart (Morgan & Saffran, 1995). Music displays contrast patterns among pitch intervals in speech and among durations of its successive vowels (Patel, 2008). Rhythms of speech and of music are correlated, especially duration contrasts between successive vowels and notes (Patel et al., 2006). Repeated entrainment to musical rhythms hones precise auditory timing, which in turn enhances the use of durational cues for speech segmentation and consonant discrimination that underpin PA (Tierney & Kraus, 2014).

Melody is the flow of pitched sounds across time. People with strong melody perception skills tend to exhibit enhanced linguistic abilities, particularly in recognising the expressive features of speech that convey emotions. According to the review by Pino et al. (2023), melody is described as the pattern of pitched sounds unfolding over time, and its perception relies on auditory discrimination and pattern recognition, abilities that are also crucial for language processing. Furthermore, pitch variability in music is correlated with English stress patterns (Patel, 2008). Specifically, the variability in interval size between consecutive pitches in speech melody parallels that of music, showing the connection between them (Patel et al., 2006). Pitch sensitivity strengthens the spectral processing of vowels and consonants. Strengthening pitch sensitivity would, therefore, improve perception of vocalic and consonantal sounds, with consequent improvement in PA (Loui et al., 2011).

We respond to music via sounds and via symbols (Heller & Campbell, 1977). Each has two levels: macro-structure and micro-structure. Macro-structure is the large formal structure, such as harmonic and motivic development. By contrast, micro-structure comprises the interpretive components (e.g., changes of stress within phrases and timbre within notes). Learning music structure is like learning language rules (Heller & Campbell, 1976), specifically like how we learn to speak in Shinichi Suzuki's "mother tongue approach" to music learning (Kendall, 1986). Humans have innate abilities to learn language and musical instruments (Kendall, 1986), learning each via listening and imitation (theory of informal music learning, Music Futures project; Green, 2008).

Musical training has been shown to enhance pitch perception across both musical and linguistic contexts (Tillmann, 2014). Likewise, a tone-language background can influence relative pitch processing, particularly in the perception of pitch intervals and melodic contours (Tillmann, 2014).

3. Music-to-speech transfer

Consider how a person's music and language skills are related. A young student Sarah played the violin (a pitched instrument) for years. Her strong aural skills let her easily distinguish among pitches and among rhythms (acoustic sensitivity). When she begins learning Mandarin, a tone language, she is better than peers without musical training at hearing and distinguishing lexical tones and stress patterns, which supports tasks like tone identification and phonological awareness (PA) in the new language. In this way, her music skills aid her language learning (*transfer*) (e.g., Kraus & Chandrasekaran, 2010; Wong et al., 2007).

Many studies show that music and speech mutually reinforce one another. As music and speech both convey information via pitch, timing, and timbre, skillfully processing them in music can aid processing them in speech (Kraus & Chandrasekaran, 2010). Notably, individuals with intensive music experience show stronger auditory encoding of pitch information in speech sounds (Wong et al., 2007). Hence, learning pitched instruments like the piano and violin can help identify tones and learning of words in a foreign language (Choi et al., 2023). Study by Patscheke et al. (2018) supports transfer from music training to PA, with evidence favouring pitch-based interventions for improving PA at the word-level in preschoolers.

The Precise Auditory Timing Hypothesis (PATH) proposed by Tierney and Kraus (2014) explains how musical training can enhance PA by sharpening the brain's temporal precision in processing sound. PATH posits that rhythm-based music practice, especially auditory-motor entrainment, the coordination of movement with auditory cues, strengthens the neural mechanisms that represent timing with high accuracy.

In music, rhythm is the patterned arrangement of sounds and silences over time. Mastering rhythm requires performers to align their movements (for example, moving a violin bow) with precise sound onsets, which is an ability termed auditory-motor entrainment. In speech, timing-based cues such as voice onset time and the durations of formant transitions support the differentiation of consonants (Elmer et al., 2013; Lin & Wang, 2011; Zuk et al., 2013). Temporal cues also facilitate word and phrase boundaries in continuous speech (Mohajer & Hu, 2003). Because accurate consonant discrimination and effective speech segmentation supports the development of PA (McBride-Chang, 1995), the PATH argues that rhythm-focused musical training, particularly that which engages auditory-motor entrainment, enhances sensitivity to durational cues in speech, thereby improving PA.

4. Speech-to-music transfer

Conversely, linguistic skills can enhance musical skills (Bidelman et al., 2013; Wong et al., 2012). People who speak tonal languages hear musical pitch better. Mandarin (a tonal language) speakers outperformed non-tonal language speakers on pitch discrimination accuracy, pitch sensitivity, interval distance sensitivity (Zhang et al., 2020), and the Musical Ear Test (Chen et al., 2016). Similarly, Cantonese (a tonal language) speakers outperformed their English and French-speaking (non-tonal languages) counterparts in musical pitch perception (Wong et al., 2012). Furthermore, even after controlling for non-verbal intelligence and working memory, Cantonese listeners surpass English non-musicians in self-designed tasks assessing musical pitch memory and discrimination (Bidelman et al., 2013). Collectively, these studies suggest that learning a tonal language sharpens sensitivity to musical pitch. Furthermore, a recent study found that Cantonese speakers outperformed Mandarin speakers in musical pitch discrimination, suggesting that learning a more complex tone language facilitates musical pitch discrimination to a larger extent (Choi & Chan, 2025).

In Rivers' study (1981), students who discriminated sounds, especially tones (aural ability), more accurately than others often better discerned and replicated a language's tonal nuances and subtle sounds (e.g., pronunciation; Rivers, 1981). During such discernment (i.e., listening), a student selected cues from speech sounds that constitute words. Later, the student chose and properly represented the appropriate phonetic elements and nuances to pronounce accurately and to express word meaning effectively (Lundsteen, 1971), especially when learning a second language (Failoni, 1993).

5. Absolute pitch ability, music training, and speech

People who can identify the pitch of any tone without a reference (absolute pitch, AP) outperform others at pitch processing and encoding (Hutka & Alain, 2015). Indeed, their brains' planum temporalia (which processes higher-order auditory and speech perception) differ morphologically and neurophysiologically (Oechslin et al., 2010).

Researchers estimate that 0.01% to 1% of people have AP, and that it is unevenly distributed (Lenhoff et al., 2001; Levitin & Rogers, 2005), being more common among tonal language speakers (Deutsch & Henthorn, 2004). Such speakers naturally acquire it as children before they turn six (Baharloo et al., 1998). During this critical period for speech acquisition (Deutsch & Henthorn, 2004), the pitches that they hear stimulate new neural circuits to perceive pitch or fine-tune old ones (Deutsch et al., 2006). According to the tone language hypothesis, languages that rely on tones to convey meaning (tonal languages, like many East Asian ones) offer grist for developing AP, perhaps via a mechanism analogous to early-onset musical training (Loui, 2018).

6. Mechanistic pathways and preregistered predictions

MSIP targets core musical skills—pitch and rhythm—that map onto auditory cues foundational to PA. In this study, tonal-rhythmic sensitivity refers to the ability to perceive and process pitch- and time-based information in a musical context. It is measured using the abbreviated Primary Measures of Music Audiation (aPMMA-T/R) and an Absolute Pitch (AP) test. Phonological awareness (PA) refers to the metalinguistic ability to consciously attend to and manipulate the sound structure of spoken language, encompassing syllables, phonemes, and suprasegmental prosody such as stress and intonation (Anthony & Lonigan, 2004; Goswami & Bryant, 1990). Because kindergarteners may have limited lexical knowledge, we focus on suprasegmental PA measures.

First, enhanced precision in tracking and labeling pitches (indexed by gains on aPMMA-T and APT) aids extraction of fundamental frequency (f0) contours; accordingly, larger tonal gains are predicted to yield greater improvements on PA (indexed by gains on EPA) that rely on pitch-based prosodic or tonal cues (e.g., tone or stress discrimination).

Second, along the temporal-rhythmic pathway, rhythm training should heighten sensitivity to temporal-envelope modulations, metrical/beat entrainment, and duration contrasts, thereby supporting speech segmentation, syllabification, and detection of within-word boundaries; thus, gains on aPMMA-R are expected to preferentially predict improvements on timing-reliant EPA measures.

Third, a prosodic integration pathway posits that coordinated pitch-and-rhythm activities shift weighting toward integrative use of f0, duration, and intensity, improving robustness of segmentation and manipulation across EPA tasks; therefore, combined gains on aPMMA-T and aPMMA-R should explain more variance in composite PA outcomes than either alone, particularly for tasks requiring cue integration.

Preregistered mediation hypotheses are: MSIP → ΔaPMMA-T/ΔAPT → ΔPA; MSIP → ΔaPMMA-R → ΔEPA; and MSIP → [ΔaPMMA-T + ΔaPMMA-R] → ΔEPA (integrative pathway). We will use multilevel mediation models to test whether music-induced gains in tonal and rhythmic sensitivity mediate improvements in PA (total and subdomains: rhyme, syllable segmentation, blending; optionally phoneme-level tasks), and we will examine cue specificity by mapping pitch gains to pitch-reliant subtasks and rhythm gains to timing-reliant subtasks, including an interaction term (ΔaPMMA-T × ΔaPMMA-R) to probe the integration pathway.

7. Cultural and educational context: Hong Kong

In Hong Kong, many children grow up speaking Cantonese, a tone language, but start learning English in kindergarten. There, they join music, drama, dance, and art activities (Curriculum Development Council, 2017). Like many kindergarten curricula across the globe, Hong Kong's blends school subject into a whole education (Vlah et al., 2019).

Pre-service kindergarten teachers receive an interdisciplinary and generalist teacher training in many areas, like music, but often lack expertise in them (Kong, 2025a, 2025b; Kong & Xiong, 2025). Some courses mix music with other arts (e.g., visual, dance, drama) or include it in broader arts courses (Lau & Grieshaber, 2018), yielding minimal musicianship training (Carroll & Harris, 2023). Hence, many kindergarten teachers teach via repetition during teacher-centred instruction (Yeung & Bautista, 2025) with few activities that foster their students' creativity (Cheung, 2017).

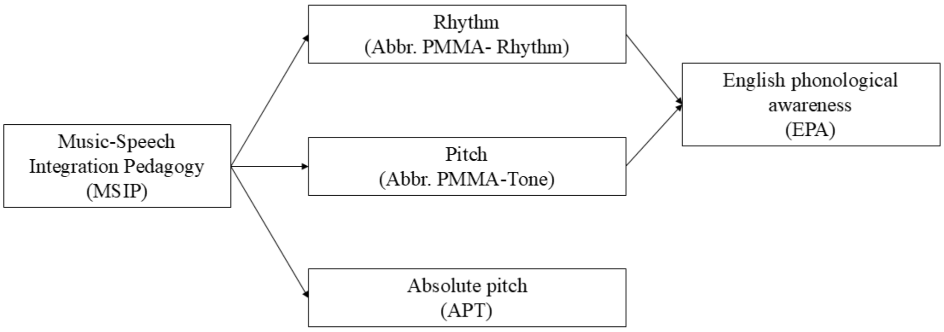
8. Limitations of prior music interventions and design rationale for the MSIP

Despite the close kinship between music and speech, intervention studies on their transfer showed mixed results, in part because both music training and outcomes varied widely (see review of 46 studies by Dumont et al, 2017) and were seldom designed to teach music via language-related methods. In the review of 46 studies, Dumont et al. (2017) highlighted substantial heterogeneity, and many studies use broad, generic music activities (listening, singing, instrumental play, movement, improvisation/composition, Orff/Kodály) rather than interventions that explicitly manipulate speech-relevant acoustic-phonological cues. A meta-analysis of controlled training studies by Gordon et al. (2015) found a small but significant aggregate benefit of music training on PA ($d \approx 0.20$), with rhyming gains increasing with greater training dosage and an estimated threshold of roughly 40 hours for reliable improvements, but no significant overall effect on reading fluency. Their effects on PA and their effects on auditory and phonological skills were mixed.

Consistent with these heterogeneous findings, a reanalysis of meta-analytic data showed small but reliable far-transfer effects of musical training when near/far transfer and control designs were properly matched (g ≈ 0.20–0.24), suggesting that mixed outcomes partly reflect methodological imbalances (Bigand & Tillmann, 2022). Kragness et al. (2021) studied children aged 5 to 10 taking music lessons for five years but did not specify the type of training, which did not affect musical ability (Kragness et al., 2021). Moreno et al.'s (2009) showed that children receiving the musical training intervention, which was based on a combination of the Kodály, Orff, and Wuytack methodologies, distinguished pitches better than other children but did not assess their musical ability, which might affect PA (Choi, 2022). Furthermore, piano training uniquely enhanced cortical responses to pitch changes in both music and speech (Nan et al., 2018).

This registered report created and tests MSIP, which capitalises on their commonalities to improve tonal-rhythmic sensitivity, AP ability, and PA. Specifically, rhythm and stress share similar acoustic cues and pitch and tone share common cues, so we hypothesise that rhythmic and pitch sensitivity in music (tested by aPMMA-T/R and APT) aid PA (tested by EPA). This report will delineate how we assess MSIP's effect on acoustic sensitivity in music pitch and rhythm, and on PA (see fig. 1).

(fig. 1) Conceptual framework.



9. Intervention: MSIP

MSIP uses music and speech's shared processes to teach effectively. Our intervention will consist of twelve 30-minute sessions of MSIP, offered outside school hours. For consistent content delivery, our study has one tutor. The tutor will be a registered early childhood education teacher in Hong Kong with an ABRSM Grade Six Piano certificate. After the tutor receives MSIP training from our research team, she/he will conduct the music training based on the provided lesson plans. The following sections will explain the theory and teaching methods.

10. Pedagogical framework

MSIP uses three parallel learning activities: (1) listening, (2) tonal-rhythmic training (syllable and note), and (3) imitation and repetition. The rhythm training activities are (1) eurythmic activity, (2) body percussion and percussion instrument playing, and (3) percussion playing as ostinato accompaniment and ensemble. The tone training activities are (1) moving with the pitch of melody contour, (2) singing melody in Solfège, (3) reproducing melody or notes on an electronic organ, and (4) listening to melody singing in syllables and identifying its Solfège. During these activities, participating children will learn via imitation and repetition. The musical materials were based on the white-key diatonic scale (C–D–E–F–G–A–B) in twelve-tone equal temperament, with A4 tuned to 440 Hz. They comprised Western children's rhymes and several original short melodies composed by the research team.

MSIP's songs map prosodic cues onto musical structure. Syllabic melodies in Solfège emphasise pitch height and contour (f0), while coordinated movement and dynamics mark metrical accents (duration/intensity). Short melodic variations in the final bar train flexible contour tracking. We aim to improve children's generalised sensitivity to stress cues and pitch, not memorisation. Classroom materials exclude the musical examples and words in the aPMMA and AP tests to reduce assessment confounds.

To link speech rhythm to musical meter, each session includes 1–2 minutes of spoken-word chants synchronised with body percussion (e.g., clap–pat–clap–rest at 80–100 bpm). Chants instantiate trochaic (STRONG-weak) and iambic (weak-STRONG) feet using nonsense syllables ("TA-ta," "ta-TA"). Louder/longer gestures reinforce metrical accents. Call-and-response formats (tutor leads, children echo) foster temporal prediction and cue integration without requiring word knowledge.

10.1. Listening as a core training activity

Learning both music and language requires attentive listening, and the MSIP aims to strengthen this skill through purposeful listening activities (Dunbar-Hall, 1991). Within MSIP, listening aligns with Kodály's principle of active aural awareness, which cultivates focused attention to pitch and rhythmic detail through repeated, active listening rather than visual or theoretical interpretation (Houlahan & Tacka, 2014). From the Orff's principle, MSIP incorporates listening as part of experiential engagement—children listen to sounds as they move, imitate, and explore musical ideas through play (Orff et al., 1980). Together, these approaches guide learners to connect auditory perception with bodily and expressive response.

MSIP further uses music-and-movement tasks to help children internalise rhythm and pitch. Coordinated actions such as clapping, stepping, or tracing melodic contours help children perceive and produce rhythmic and intonational features in both song and speech (Lado, 1964, p. 80).

10.2. Tonal-rhythmic training (Syllable and note)

In the tonal–rhythmic training component, MSIP highlights perceptual parallels between the organisation of sound in music and in language. A syllable often corresponds to a single note—although a syllable may extend over several notes in melismatic singing (Lerdahl, 2013). Typically, the consonant–vowel structure of a syllable resembles a note's attack and sustained pitch. Syllables combine into words and phrases, much as notes combine into motives and phrases. While Lerdahl (2013) situates these correspondences within his generative model of musical syntax, here the focus is specifically on how these cross-modal parallels can strengthen children's perception of rhythm, pitch, and prosody within the MSIP framework.

10.3. Imitation and repetition

Imitation and repetition are central mechanisms in both music learning and language acquisition, particularly in early childhood. According to Särkämö et al. (2013), children are naturally drawn to music and begin to imitate musical sounds and gestures from a very young age. As early as six months, infants begin to babble rhythmically and move in synchrony with musical beats—demonstrating the embeddedness of imitation in musical engagement.

These early behaviours are not merely spontaneous but serve critical developmental functions. Music provides a multisensory platform where children can reproduce pitch, rhythm, and melodic contours, often through singing, clapping, or movement. Through repetition, these auditory-motor patterns become internalised, supporting the development of auditory memory, attention, and expressive communication (Särkämö et al., 2013).

This view aligns with the OPERA and PATH theoretical framework. The OPERA hypothesis (Patel, 2011) posits that musical activities—especially those involving imitation and repetition—can drive adaptive plasticity in brain networks shared with language, particularly when they are emotionally engaging, precise, and repeated. Similarly, the PATH (Tierney & Kraus, 2014) suggests that rhythm-based practice, reinforced through repetition and imitation, enhances temporal precision in auditory processing, which is foundational for speech segmentation and phonological awareness. These frameworks support the idea that repetitive musical engagement strengthens shared auditory-motor pathways, enabling transfer from music to speech processing skills.

Furthermore, imitation and repetition are deeply intertwined with speech development. Infant-directed speech (or motherese) naturally incorporates melodic and rhythmic exaggeration, inviting infants to imitate prosodic features that are foundational for later phonological awareness. As children engage in repetitive musical activities—such as singing familiar songs or echoing rhythmic patterns—they not only refine their musical skills but also strengthen neural circuits shared by music and speech processing.

10.4. Musical instruments

The human brain's AP processing handles voices differently than it does musical instruments and pure tones (Gao & Oxenham, 2022). Specifically, pitch judgments for voices tend to be less accurate due to interference from voice-specific processing mechanisms that prioritise speech-related cues over fine-grained pitch information—a phenomenon known as the voice disadvantage effect. To avoid this complication and ensure that children's pitch perception is not confounded by vocal timbre processing, we will use Yamaha Electone StageA ELS-02c and StageA ELB-01 (tuning: A4 = 440 Hz) for the class activities and materials. In class, the students will play music on a Yamaha Electone StageA ELB 01.

In addition, the cabasa was chosen as the percussion instrument in this study because it is easy for kindergarteners to handle and can produce both short and sustained ('long') sounds with different playing techniques (e.g., shaking, twisting, or striking), offering varied sound textures that help children feel and reproduce rhythmic duration.

11. MSIP lesson plan protocol: Structure, progression, and dose

Each 30-minute MSIP session follows a set sequence to maximise aural exposure and opportunities for immediate feedback in small groups (n = 5). They begin with coordinated movement and fixed-do Solfège (≈ 5 min), followed by rhythmic imitation using body percussion, rhythmic syllables, and cabasa (≈ 5 min). A Solfège "fast response" echoing block (≈ 5 min) precedes song singing in Solfège (≈ 5 min) and playing instruments like keyboards or desk bells (≈ 5 min). Sessions end with music-and-story pitch games (≈ 3 min) and a short reflection (≈ 1–2 min). This routine helps young children know what to expect while ensuring a high number of short, repeated learning and practice chances in each component.

We measure the minimum frequency of musical exercises playing in the programme to document expected per-child exposure. Each Solfège "fast response" block typically comprises 8 bars presented for 2–3 repetitions (≈ 16–24 tonal echo trials). Three Solfège "fast response" block will be introduced in each session. Instrument blocks add ≈ 8–12 reproduction trials as children progress from air-play to keyboard execution. Rhythmic imitation—including body percussion, rhythmic syllables, and cabasa—yields ≈ 12–16 bar-level echoes per session. Overall, each child completes ≈ 56–84 tonal trials and ≈ 12–16 rhythmic trials per session, totaling more than 672–1,008 tonal and 144–192 rhythmic trials over 12 sessions, in addition to integrated prosody mapping embedded in song and story activities. This high-density, imitation-based practice aligns with prior research showing short-term auditory gains for kindergarteners (Moreno et al., 2011). Detailed lesson plans are available upon request.

Manual delivery of training and structured monitoring support implementation fidelity. The tutor completes a per-session checklist for each lesson block with actual counts of bars and repetition targets. Twenty percent of sessions are audio/video recorded for independent raters to check adherence to the lesson plans, aiming for at least 85% adherence. We will record children's attendance at each session and analyse these data to determine how well children respond to the programme, including how variation in each child's attendance relates to outcomes.

12. Study design and methods

12.1. Trial design

This study is a prospective, parallel-group, two-armed, randomised controlled trial. It has a pre-test, a post-test, and a delayed post-test, along with an intention-to-treat primary analysis.

12.2. Sample size

Yashaswini and Maruthy's (2020) music training study showed an effect size on categorical perception of 0.5, and studies of Hong Kong kindergarteners show little attrition (often < 5%). Hence, our study with an expected effect size of 0.5, 10% attrition, 80% power, and $\alpha = .05$ requires 28 participants (14 participants per group), according to G*Power software. To account for potential attrition, 40 participants will be recruited in this study.

12.3. Participants

Our non-probability, volunteer convenience sample will be Cantonese-speaking children aged 4 – 5 years in Hong Kong. We will recruit them via flyers in kindergartens and online channels like Facebook advertisement. Interested parents will complete a brief screening survey about their child's age, home language, music training, and special educational needs. Participants must be 4 – 5 years old, speak Cantonese as their first language, be enrolled in a Hong Kong kindergarten, and have their parent's consent to participate in this study and complete all assessments. We will exclude those receiving structured music training or with special educational needs (according to their parents). Then, eligible children will complete a baseline assessment, and their parents will complete a survey about demographics (parents' educations, incomes, child's age and gender) and their child's kindergarten type (non-profit local vs private independent). Then, we will randomly assign half the children to the experimental group and half to the control group via a computer-generated sequence with permuted blocks of variable size, stratified by age (4 vs 5 years) and gender.

13. Procedure

All 40 children will individually complete a pre-test, a post-test (immediately after the intervention), and a delayed post-test (two months later). Each test comprises the abbreviated Primary Measures of Music Audiation (aPMMA), Absolute Pitch Test (APT) and English Phonological Awareness (EPA). A research assistant with early childhood education teacher training will administer each test to each child.

Short, focused aural training has yielded significant behavioural and neural changes in preschoolers' auditory and language-related processing (Moreno et al., 2011). Following this model, groups of three to five children in the intervention condition will receive twelve 30-minute sessions of music training, totaling six hours. The 12×30-minute small group format yields several hundred high-fidelity stimulus-response cycles per child across tonal and rhythmic domains (See "MSIP Lesson Plan Protocol: Structure, Progression, and Dose"). For consistent content delivery, one tutor will do all the music training, closely following our lesson plans. Children in the control condition will receive their usual kindergarten music instruction. We will quantify each child's actual exposure by tracking attendance.

14. Allocation concealment and blinding

We will maximally conceal allocation of children to the experimental or control conditions with sequentially numbered, opaque, sealed envelopes opened only after baseline completion. Outcome assessors will be blinded to group assignment when feasible. We will use masked group codes for the primary analyses.

15. Outcome measures: Music

15.1. aPMMA

We assess these children's tonal and rhythmic sensitivities with an objective, short (ten minute), child-friendly test, the aPMMA (Gordon, 1979). Extracted from the full PMMA, aPMMA has practice examples and 30 tape-recorded short musical phrases: 15 Tonal (aPMMA-T) and 15 Rhythmic (aPMMA-R). The Tonal subtest has short tonal patterns without rhythmic variations. Conversely, the Rhythmic subtest has short rhythm patterns with all notes at the same pitch. Each child receives an answer sheet with yes (✓) and no (×) symbols for each of the 30 items. For each item, a child hears recordings of two tonal or rhythmic patterns. If they sound the same, the child circles the yes symbol. Otherwise, the child circles the no symbol.

15.2. Absolute Pitch Test (APT)

Our study investigates the relationship between musical pitch sensitivity and PA in English. Including AP as a variable provides an index of individual differences in auditory encoding precision, which may influence how children perceive and process linguistic prosody. In this context, AP serves as a measure of fine-grained pitch acuity that facilitates sensitivity to tonal and prosodic cues in speech, thereby potentially interacting with PA.

Rather than assessing AP, we will assess how well each child identifies AP-like pitch within a single octave (Recognising the correct pitch chroma is key to AP) via a test based on Oechslin et al. (2010). Aligned with the fixed-do instructional materials, a child will hear a clarinet note (C4, D4, E4, F4, F#4, G4, A4, or B4) for one second (tuning: A4 = 440 Hz). Then, the child will have four seconds to identify its tone on the Solfège scale (do, re, mi, fa, fi, sol, la, ti) while listening to brown noise. They will hear a sequence of 13 notes in a set order, with each note occurring three times. We produce all notes with the Yamaha Electone StageA ELS-02c (tuning: A4 = 440 Hz). *Accuracy* is the ratio of correct answers over 13 total notes.

16. Outcome measures: Speech

16.1. English Phonological Awareness (EPA)

We use syllable-deletion and onset-phoneme–deletion tasks adapted from prior work (Choi et al., 2025). The syllable-deletion task includes 6 test items: on each, children will hear an English word (e.g., spider /ˈspaɪdɚ/) and then repeated it with one syllable removed (e.g., say /ˈspaɪdɚ/ without /ɚ/). The onset-phoneme–deletion task comprised 17 test items: on each, children will hear a word (e.g., Mike /maɪk/) and then produce it without a specified phoneme (e.g., say /maɪk/ without /k/). Both tasks will begin with three practice trials with feedback. Each correct test response will earn 1 point; scores will be summed and divided by 23 to yield percent accuracy. The task showed high internal consistency (Cronbach's α = .93) (Choi et al., 2025).

17. Statistical analysis

We will summarise baseline characteristics by group with means/SDs, medians/IQRs, and proportions and report standardised mean differences (SMDs; an SMD > 0.20 is notable). To accurately test the effect of this intervention, we will address the following issues with specific statistics strategies: (a) missing data with Markov Chain Monte Carlo multiple imputation (Peugh & Enders, 2004), (b) selection bias with five statistical matching methods (propensity score matching, Mahalanobis distance matching, coarsened exact matching, Mahalanobis frontier, L1 frontier; Stuart, 2010), (c) similarities versus differences of responses across trials, pairs, trial types, and students with multilevel (ML) analysis (Hox et al., 2017), (d) multiple outcomes with multivariate outcome multilevel analysis (Hox et al., 2017), (e) omitted variable bias with difference-in-differences analyses (Bertrand et al., 2004), (f) indirect, ML mediation effects with multilevel M-tests (MacKinnon et al., 2004) (g) cross-level interactions (task x student) with random parameters in random effects models (Hox et al., 2017), (h) many hypotheses' false positives with the two stage linear step-up procedure (Benjamini et al., 2006), (i) compare effect sizes with Lagrange multiplier tests (Bertsekas, 2014), and (j) consistency of results across data sets (robustness) with separate multilevel, single outcome models; analyses of data subsets; and analyses of original (not estimated) data (Kennedy, 2008; for details, see Chiu & Lehmann-Willenbrock, 2016).

17.1. Explanatory model

We will model students' outcomes with a multivariate outcome, multilevel difference-in-differences analysis, starting with a variance components model to test for significant differences at each level: test question and student (Hox et al., 2017).

$$\mathbf{Test}_{yijkl} = \hat{a}_y + e_{yijkl} + f_{yjkl} + g_{ykl} + h_{yl} + \hat{a}_s\mathbf{Post}_y + \hat{a}_t\mathbf{Grouo}_y + \hat{a}_u\mathbf{Post*Group}_y + \hat{a}_v\mathbf{Delayed\_Post}_y + \hat{a}_w\mathbf{Delayed\_Post*Group}_y \quad (1)$$

In the vector **Test**yij, outcome y (aPMMA test question, APT question, EPA question) of response to trial *i* of pair *j* in trial type *k* by student *l* had grand mean intercept $\hat{a}_y$, with unexplained components (*residuals*) at the trial-, pair-, trial type-, and student-levels (e_{yij}, f_{yj}). Then, we will enter the vectors **Post** (post-test vs pre-test), **Delayed_Post** (delayed post-test vs. pre-test), **Group** (experimental condition vs control), and their interactions **Post*Group** and **Delayed_Post*Group**—which test for significant differences in learning across groups.

Next, we will test the extent to which demographics accounted for or mediated learning across groups with a multilevel structural equation model.

$$\mathbf{Test}_{yijkl} = \hat{a}_y + e_{yijkl} + f_{yjkl} + g_{ykl} + h_{yl} + \hat{a}_s\mathbf{Post}_y + \hat{a}_t\mathbf{Group}_y + \hat{a}_u\mathbf{Post*Group}_y + \hat{a}_v\mathbf{Delayed\_Post}_y + \hat{a}_w\mathbf{Delayed\_Post*Grouo}_y + \hat{a}_{yx}\mathbf{Demographics}_{yl} + \hat{a}_{yzjkl}\mathbf{Interactions}_{yijkl} \quad (2)$$

Lastly, we will test for Interactions among significant explanatory variables.

18. Insights from pilot study: PMMA test adaptation

In our May 2024 pilot study, we used the full PMMA (Gordon, 1979) and found that it took too long for children aged 4 – 5 to complete—over 40 minutes. To reduce its length, we applied reliability analysis to reduce the number of items measuring pitch and rhythm from 40 to 15.

Internal consistencies were low for the original 40 tone items (Cronbach's α = .58) and 40 Rhythm items (α = .68). So, we removed the items with poor item-total correlations below 0.3. The remaining items had high internal consistency (15 Tone items: α = .91; 15 Rhythm items: α = .90). Kindergarteners can complete these 30 items in 10 minutes without feeling bored or tired.

Also, the original answer sheet with two identical and two different faces for each item confused some children. So, we replaced them with a simplified format of yes (✓) and no (×) symbols. Kindergarteners readily understood this new format.

19. Significance of the study

We will test whether MSIP capitalises on the connections between music and language skills to help kindergarteners learn these skills. Our findings will inform interventions to improve music skills of tonal-rhythmic sensitivity and AP ability (near transfer) and PA (far transfer). Furthermore, the results may tell us more about the mechanisms underlying the resemblances in music and speech perception or learning. These results may also inform speech and language therapy, especially interventions for individuals with speech disorders or attention deficits.

Furthermore, MSIP is simple and costs little, using only desk bells and cabasas. As teachers only need basic musical playing skills (not extensive musical training) to use MSIP, they can do it in preschool classes. By offering a framework that blends music and speech activities into the current curriculum, our study will inform young children's development of acoustic sensitivity in preschools.

20. Limitations

This study has three major limitations: participants, study design, and single site. First, our study only includes young, Cantonese-speaking children (4 – 5 years old). Such young children have high auditory plasticity that can help them learn pitch and rhythm (Saffran et al., 1996). Also, speaking a tone language can shape their baseline pitch processing and cue weighting, which can change how they learn English stress patterns. Hence, our findings might not generalise to older children or those who only speak a non-tonal language(s).

Second, our multi-component, short-duration design (12 sessions) integrates MSIP's tonal echo-singing, rhythmic imitation with body percussion and cabasa, brief chant-based speech-beat integration, and instrument reproduction. Without isolating the unique contribution of each MSIP activity, we cannot determine which one(s) drives enhancement of each music or language skill. Future dismantling or factorial designs can identify MSIP's active ingredients and their interactions.

Third, we only test MSIP at one site. Despite our manualisation and fidelity checks, one tutor using MSIP to teach groups of children at one site cannot account for differences across classrooms, schools, or cultures.

Ethical consideration

The first author's university approved the ethics of our study. Subsequently, we will seek the consent of participating children and parents, de-identify the data, report them in aggregate, and store them away from our analytic results and manuscript.

Funding

This study was funded by the Departmental Research Grant from the Department of Early Childhood Education at the Education University of Hong Kong (Project ID: DRG2022-23/004).

21. References

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early primary school children. *Journal of Educational Psychology*, *96*(1), 43–55.

Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., & Freimer, N. B. (1998). Absolute pitch: An approach for identification of genetic and nongenetic components. *American Journal of Human Genetics*, *62*(2), 224–231. https://doi.org/10.1086/301704

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika, 93*(3), 491–507.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, *119*(1), 249–275.

Bertsekas, D. P. (2014). *Constrained optimisation and Lagrange multiplier methods*. Academic Press.

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PloS one*, *8*(4), Article e60676. https://doi.org/10.1371/journal.pone.0060676

Bigand, E., & Tillmann, B. (2022). Near and far transfer: Is music special? *Memory & Cognition*, *50*(2), 339–347. https://doi.org/10.3758/s13421-021-01226-6

Carroll, C., & Harris, J. (2023). 'Because I'm not musical': A critical case study of music education training for pre-service generalist primary teachers in Australia. *British Journal of Music Education*, *40*(2), 271–286. https://doi.org/10.1017/S0265051722000274

Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience*, *31*(6), 751–760. https://doi.org/10.1080/23273798.2016.1156715

Cheung, R. H. P. (2017). Teacher-directed versus child-centered: The challenge of promoting creativity in Chinese preschool classrooms. *Pedagogy, Culture & Society*, *25*(1), 73–86. https://doi.org/10.1080/14681366.2016.1217253

Chiu, M. M., & Lehmann-Willenbrock, N. (2016). Statistical discourse analysis: Modeling sequences of individual behaviors during group interactions across time. *Group Dynamics: Theory, Research, and Practice*, *20*(3), 242–258. https://doi.org/10.1037/gdn0000048

Choi, W. (2022). What is "Music" in music-to-language transfer? Musical ability but not musicianship supports Cantonese listeners' English stress perception. *Journal of Speech, Language, and Hearing Research*, *65*(11), 1–13. https://doi.org/10.1044/2022_JSLHR-22-00175

Choi, W., & Chan, L. Y. (2025). The comparative effects of Cantonese and Mandarin tone language backgrounds on musical pitch perception. *Scientific Reports*, *15*, Article 31558. https://doi.org/10.1038/s41598-025-17171-2

Choi, W., To, C. Y., & Cheng, R. (2023). The choice of musical instrument matters: Effect of pitched but not unpitched musicianship on tone identification and word learning. *Applied Psycholinguistics*, *44*(5), 844-857. https://doi.org/10.1017/S0142716423000358

Choi, W., Lai, V. K. W., Kong, S.-H., & Bautista, A. (2025). Examining the cognitive and perceptual perspectives of music-to-language transfer: A study of Cantonese–English bilingual children. *Journal of Experimental Child Psychology*, *249*, Article 106069. https://doi.org/10.1016/j.jecp.2024.106069

Creel, S. C., Obiri-Yeboah, M., & Rose, S. (2023). Language-to-music transfer effects depend on the tone language: Akan vs. East Asian tone languages. *Memory & Cognition*. *51*. 1624–1639. https://doi.org/10.3758/s13421-023-01416-4

Curriculum Development Council. (2017). *Kindergarten education curriculum guide*. Education Bureau.

Deutsch, D., & Henthorn, T. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception*, *21*(3), 339–356. https://doi.org/10.1525/mp.2004.21.3.339

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America*, *119*(2), 719–722. https://doi.org/10.1121/1.2151799

Dumont, E., Syurina, E. V., Feron, F. J., & van Hooren, S. (2017). Music interventions and child development: A critical review and further directions. *Frontiers in Psychology*, *8*, 1694.

Dunbar-Hall, P. (1991). Music and language: commonalities in Semiotics, syllabus, and classroom teaching. *British Journal of Music Education*, *8*(1), 65–72. https://doi.org/10.1017/S026505170000807X

Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex*, *49*(10), 2812–2821. https://doi.org/10.1016/j.cortex.2013.03.007

Failoni, J. W. (1993). Music as a means to enhance cultural awareness and literacy in the foreign language classroom. *Mid-Atlantic Journal of Foreign Language Pedagogy*, *1*, 97–108.

Gao, Z., & Oxenham, A. J. (2022). Voice disadvantage effects in absolute and relative pitch judgments. *The Journal of the Acoustical Society of America*, *151*(4), 2414–2428. https://doi.org/10.1121/10.0010123

Gordon, E. E. (1979). *Primary measures of music audiation*. GIA.

Gordon, R. L., Fehd, H. M., & McCandliss, B. D. (2015). Does music training enhance literacy skills? A Meta-Analysis. *Frontiers in Psychology*, *6*, 01777. https://doi.org/10.3389/fpsyg.2015.01777

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Lawrence Erlbaum.

Green, L. (2008). *Music, informal learning and the school: A new classroom pedagogy*. Ashgate.

Heller, J., & Campbell, W. C. (1976). Models for language and intellect in music research. *Music Education for Tomorrow's Society*, 40–49.

Heller, J., & Campbell, W. C. (1977). The relationship between the interpretive element in music and the acoustic microstructure. *Bulletin of the Council for Research in Music Education*, *50.* 29–33. http://www.jstor.org/stable/40317442

Houlahan, M., & Tacka, P. (2014). *Kodály in the kindergarten classroom: Developing the creative brain in the 21st century*. Oxford University Press.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis*. Routledge.

Hutka, S. A., & Alain, C. (2015). The effects of absolute pitch and tone language on pitch processing and encoding in musicians. *Music Perception*, *32*(4), 344–354. https://doi.org/10.1525/mp.2015.32.4.344

Kendall, J. (1986). Suzuki's Mother Tongue Method. *Music Educators Journal*, *72*(6), 47–50. https://doi.org/10.2307/3401277

Kennedy, P. (2008). *Guide to econometrics*. Wiley-Blackwell.

Kong, S. H. (2025a). Music education training for kindergarten teachers: a workshop integrating Shulman's pedagogical content knowledge and pedagogical reasoning and action model. *Journal of Early Childhood Teacher Education*, *46*(1), 85–102. https://doi.org/10.1080/10901027.2024.2410399

Kong, S.-H. (2025b). Confidence in music teaching: The impact of instrumental training, teaching experience, and pedagogical skills on early childhood educators. *International Journal of Early Childhood*. https://doi.org/10.1007/s13158-025-00447-z

Kong, S.-H., & Xiong, X. (2025). Pre-Service Kindergarten Teachers' Confidence and Beliefs in Music Education: A Study in the Chinese Context. *Education Sciences*, *15*(6), 772. https://doi.org/10.3390/educsci15060772

Kragness, H. E., Swaminathan, S., Cirelli, L. K., & Schellenberg, E. G. (2021). Individual differences in musical ability are stable over time in childhood. *Developmental Science*, *24*(4), e13081. https://doi.org/10.1111/desc.13081

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Krumhansl, C. L., & Cuddy, L. L. (2010). A theory of tonal hierarchies in music. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception* (pp. 51–87). Springer.

Lado, R. (1964). *Language teaching: A scientific approach*. McGraw-Hill.

Lau, W. C. M., & Grieshaber, S. (2018). School-based integrated curriculum: An integrated music approach in one Hong Kong kindergarten. *British Journal of Music Education*, *35*(2), 133–152. https://doi.org/10.1017/S0265051717000250

Lenhoff, H. M., Perales, O., & Hickok, G. (2001). Absolute pitch in Williams Syndrome. *Music Perception*, *18*(4), 491–503. https://doi.org/10.1525/mp.2001.18.4.491

Lerdahl, F. (2013). Musical syntax and its relation to linguistic syntax. In M. A. Arbib (Ed.), *Language, music, and the brain*. The MIT Press. https://doi.org/10.7551/mitpress/9548.003.0016

Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences*, *9*(1), 26–33. https://doi.org/10.1016/j.tics.2004.11.007

Lin, C. Y., & Wang, H. C. (2011). Automatic estimation of voice onset time for word-initial stops by applying random forest to onset detection. *The Journal of the Acoustical Society of America*, *130*(1), 514–525. https://doi.org/10.1121/1.3592233

Loui, P. (2018). Absolute pitch. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed.). Oxford University Press.

Loui, P., Li, H. C., Hohmann, A., & Schlaug, G. (2011). Enhanced connectivity in absolute pitch musicians: A model of hyperconnectivity. *Journal of Cognitive Neuroscience*, *23*(4), 1015–1026. https://doi.org/10.1162/jocn.2010.2150

Lundsteen, S. W. (1971). *Listening: Its impact on reading and the other language arts*. National Council of Teachers.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4

McBride-Chang, C. (1995). What is phonological awareness? *Journal of Educational Psychology*, *87*(2), 179–192. https://doi.org/10.1037/0022-0663.87.2.179

Mohajer, K., & Hu, Z-M. (2003). Robust speech recognition using time boundary detection. In *Multisensor, multisource information fusion: Architectures, algorithms, and applications 2003* (Vol. 5099, pp. 335–343). Society of Photo-optical Instrumentation Engineers (SPIE). https://doi.org/10.1117/12.488199

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, *22*(11), 1425–1433. https://doi/10.1177/0956797611416999

Moreno, S., Lee, Y., Janus, M., & Bialystok, E. (2015). Short-term second language and music training induces lasting functional brain changes in early childhood. *Child Development*, *86*(2), 394–406. https://doi.org/10.1111/cdev.12297

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, *19*(3), 712–723. https://doi.org/10.1093/cercor/bhn120

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*(4), 911–936. https://doi.org/10.1111/j.1467-8624.1995.tb00913.x

Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., Gabrieli, J. D. E., & Desimone, R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences*, *115*(28), E6630–E6639. https://doi.org/10.1073/pnas.1808412115

Oechslin, M. S., Meyer, M., & Jäncke, L. (2010). Absolute pitch—Functional evidence of speech-relevant auditory acuity. *Cerebral Cortex*, *20*(2), 447–455. https://doi.org/10.1093/cercor/bhp113

Orff, C., Keetman, G., & Regner, H. (1980). *Music for children*. Schott Music Corp.

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, Article 142. https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, *119*(5), 3034–3047. https://doi.org/10.1121/1.2179657

Patscheke, H., Degé, F., & Schwarzer, G. (2018). The effects of training in rhythm and pitch on phonological awareness in four- to six-year-old children. *Psychology of Music*, *47*(3), 376–391. https://doi.org/10.1177/0305735618756763

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research. *Review of Educational Research*, *74*(4), 525–556. https://www.jstor.org/stable/3515980

Pino, M. C., Giancola, M., & D'Amico, S. (2023). The association between music and language in children: A state-of-the-art review. *Children (Basel)*, *10*(5), 801. https://doi.org/10.3390/children10050801

Rivers, W. M. (1981). *Teaching foreign-language skills* (2nd ed.). University of Chicago Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Särkämö, T., Tervaniemi, M., & Huotilainen, M. (2013). Music perception and cognition: development, neural basis, and rehabilitative use of music. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(4), 441–451.

Stolarski, Ł. (2015). Pitch patterns in vocal expression of "happiness" and "sadness" in the reading aloud of prose on the basis of selected audiobooks. *Research in Language*, *13*(2), 140–161. https://doi.org/10.1515/rela-2015-0016

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*, 1–21.

Syed, M., Frank, M. C., & Roisman, G. I. (2023). Registered Reports in Child Development: Introduction to the Special Section. *Child Development*, *94*(5), 1093–1101. https://doi.org/10.1111/cdev.14003

Tierney, A., & Kraus, N. (2014). Auditory–motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience*, *8*, 949. https://doi.org/10.3389/fnhum.2014.00949

Tillmann, B. (2014). Pitch processing in music and speech. *Acoustics Australia*, 42(2), 124–130.

Vlah, N., Jančec, L., & Sabolić, M. (2019). Kindergarten teachers as models and facilitators of socialisation in the kindergarten. *Metodički Obzori*, *13(2018)1(25)*, 27–56. https://doi.org/10.32728/mo.13.1.2018.02

Wong, P., Skoe, E., Russo, N., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422.

Wong, P. C. M., Ciocca, V., Chan, A. H. D., Ha, L. Y. Y., Tan, L.-H., & Peretz, I. (2012). Effects of culture on musical pitch perception. *PloS one*, *7*(4), Article e33424. https://doi.org/10.1371/journal.pone.0033424

Yashaswini, L., & Maruthy, S. (2020). Effect of music training on categorical perception of speech and music. *Journal of Audiology & Otology*, *24*(3), 140–148. https://doi.org/10.7874/JAO.2019.00500

Yeung, J., & Bautista, A. (2025). Arts education in Hong Kong kindergartens: Provision of activities and impact of teachers' demographics. *Arts Education Policy Review*, *126*(2), 114–130. https://doi.org/10.1080/10632913.2024.2336514

Zhang, L., Xie, S., Li, Y., Shu, H., & Zhang, Y. (2020). Perception of musical melody and rhythm as influenced by native language experience. *The Journal of the Acoustical Society of America*, *147*(5), EL385–EL390. https://doi.org/10.1121/10.0001179

Zuk, J., Ozernov-Palchik, O., Kim, H., Lakshminarayanan, K., Gabrieli, J. D., Tallal, P., & Gaab, N. (2013). Enhanced syllable discrimination thresholds in musicians. *PLoS One*, *8*(12), e80546. https://doi.org/10.1371/journal.pone.0080546

Kong Siu-Hang. Assistant Professor of Music Education at The Education University of Hong Kong. His research interests include early childhood music education, teacher education, the sociology and psychology of music education, and comparative education. He is the author of two monographs, *Cultural Capital and Parental Involvement: A Comparison of Students' Music Participation between Beijing and Hong Kong* and *Soundscapes of Education: Nurturing Creativity through Music and the Arts*, and has published in various refereed journals. Drawing on his practical experience in music education, his research seeks to advance both theoretical understanding and professional practice in the field.

Chiu Ming Ming. Chair (Distinguished) Professor of Analytics and Diversity at The Education University of Hong Kong. He advises China's and Qatar's Ministries of Education. He invented Deceptive Writing Theory, statistical discourse analysis (best 50 learning science ideas –International Society of the Learning Sciences), multilevel diffusion analysis, artificial intelligence Statistician, and online detection of sexual predators. His 107 grants (US$23 million) yielded 325 publications (232 journal articles; 20,000+ citations; #8 in Education in China), 22 keynote speeches, 5 television broadcasts, 17 radio broadcasts, and 204 news articles in 22 countries. He creates automatic statistical analyses for Big Data.

Choi William. Assistant Professor of Speech-Language Pathology at The University of Hong Kong, where he also serves as the Director of the Speech and Music Perception Laboratory and the Co-Director of the Centre for Advancement in Inclusive and Special Education. He is also a Fellow of the Psychonomic Society. His research interests include music cognition, speech perception, and reading.

Bautista Alfredo. Full Professor and Associate Head of the Department of Early Childhood Education (ECE) at The Education University of Hong Kong. He has conducted educational research in countries around the world (Spain, Argentina, United Kingdom, United States, Canada, Singapore, Hong Kong/China). He currently

leads several research and development projects focusing on early childhood education curriculum, pedagogy, teacher education and professional development. Much of this prior research work has focused on music and arts education, creativity, and play. He is Editor-in-Chief of the Journal for the Study of Education and Development and serves as Associate Editor and Editorial Board Member for other peer-reviewed international journals.